# Managing Stress: The Needs of Autistic Adults in Video Calling

ANNUSKA ZOLYOMI, Microsoft Research, University of Washington
ANDREW BEGEL, Microsoft Research
JENNIFER FRANCES WALDERN, Microsoft
JOHN TANG, MIKE BARNETT, ED CUTRELL, DANIEL MCDUFF, SEAN ANDRIST, and MEREDITH RINGEL MORRIS, Microsoft Research

Video calling (VC) aims to create multi-modal, collaborative environments that are "just like being there." However, we found that autistic individuals, who exhibit atypical social and cognitive processing, may not share this goal. We interviewed autistic adults about their perceptions of VC compared to other computer-mediated communications (CMC) and face-to-face interactions. We developed a neurodiversity-sensitive model of CMC that describes how stressors such as sensory sensitivities, cognitive load, and anxiety, contribute to their preferences for CMC channels. We learned that they apply significant effort to construct coping strategies to support their sensory, cognitive, and social needs. These strategies include moderating their sensory inputs, creating mental models of conversation partners, and attempting to mask their autism by adopting neurotypical behaviors. Without effective strategies, interviewees experience more stress, have less capacity to interpret verbal and non-verbal cues, and feel less empowered to participate. Our findings reveal critical needs for autistic users. We suggest design opportunities to support their ability to comfortably use VC, and in doing so, point the way towards making VC more comfortable for all.

CCS Concepts: • **Human-centered computing → Computer supported cooperative work**; **Empirical studies in accessibility**.

Additional Key Words and Phrases: video calling, computer-mediated communication, accessibility, autism, neurodiversity

**ACM Reference Format:**
Annuska Zolyomi, Andrew Begel, Jennifer Frances Waldern, John Tang, Mike Barnett, Ed Cutrell, Daniel McDuff, Sean Andrist, and Meredith Ringel Morris. 2019. Managing Stress: The Needs of Autistic Adults in Video Calling. *Proc. ACM Hum.-Comput. Interact.* 3, CSCW, Article X (November 2019), 29 pages. https://doi.org/10.1145/TBD

## 1 INTRODUCTION

Video calling (VC) applications provide multi-modal environments that offer real-time video, audio, chat, and desktop sharing channels to create common ground for collaboration in many domains, including work, school, and personal scenarios [56]. Video calls and other computer-mediated

Authors' addresses: Annuska Zolyomi, annuska@uw.edu, Microsoft Research, University of Washington, Seattle, WA; Andrew Begel, abegel@microsoft.com, Microsoft Research, Redmond, WA; Jennifer Frances Waldern, Jennifer.Waldern@microsoft.com, Microsoft; John Tang, johntang@microsoft.com; Mike Barnett, mbarnett@microsoft.com; Ed Cutrell, cutrell@microsoft.com; Daniel McDuff, damcduff@microsoft.com; Sean Andrist, sandrist@microsoft.com; Meredith Ringel Morris, merrie@microsoft.com, Microsoft Research, Redmond, WA.


2573-0142/2019/11-ARTX $15.00
https://doi.org/10.1145/TBD

communication (CMC) tools connect people across distance, often when they would otherwise be unable to meet face-to-face (FtF). These tools can be hyper-personal—enabling the senders, receivers, communication channels, and messages to work together to bolster interpersonal relationships [53].

Some people who want to use VC are on the autism spectrum. Autism spectrum disorder (ASD) is a lifelong neurodevelopmental condition characterized by particular cognitive styles, communication behaviors, social interactions, and repetitive behaviors [3]. Recent estimates put the prevalence of autism in the USA at 1 in 42 in boys and 1 in 189 in girls [4, 5], a rate which has risen drastically over the past 30 years as awareness has increased and diagnosis processes evolve. As autistic individuals[1] reach adulthood and enter the workforce, new VC scenarios and user requirements are likely to arise to support collaboration between autistic and non-autistic (i.e., neurotypical[2]) collaborators.

During VC interactions, people's goals range from building social relationships to collaborating towards completing a group task. While attending to these meta-cognitive goals, each person has to draw upon their low-level cognitive processing capabilities to manage the multi-modal channels of audio, video, text, and images. These processing demands can be challenging for neurotypical users because of natural limitations on human processing abilities [15, 37]. As a result, researchers have examined ways to help VC users more efficiently manage meta-cognitive and low-level cognitive processing (e.g., [16, 44, 56]).

Socio-technical challenges are more intense for many autistic VC users. Due to current social norms and the design of VC systems, autistic individuals have to adapt their particular style of cognitive processing and hyper- or hypo- sensitivity to sensory inputs during a VC. Their impressions of using VC are likely impacted by how they cognitively process VC interactions. Autistic individuals report having difficulty expressing their emotions, adapting to new situations, and working through ambiguity [19, 22]; they tend to focus on details and excel at conceptualizing phenomena as systems [6]. During VC situations, their sensory sensitivities may be irritated, especially when lights prove to be too bright and microphone noises sound too loud and are distracting [19, 22]. The high-bandwidth communication channels offered by video calls may easily trigger an autistic individual's detail-focused cognitive style to devote too much time, attention, and effort trying to read others' emotions and body language, causing them to fall behind in conversational flow [22, 42]. Though VC affords "some of the intimacies of co-presence," [23, p. 68], the desire for that affordance may not be shared by autistic users.

Unfortunately, little is known about the strategies and technology decisions of autistic adults as they engage (or disengage) with VC. In this paper, we aim to fill this gap by asking

RQ1. How and why do autistic adults make use of video calling?
RQ2. What factors increase or reduce the comfort of the video calling experiences of autistic adults?
RQ3. How do these comfort-influencing factors impact the CMC channel preferences of autistic adults?
RQ4. What coping strategies do autistic adults make use of to reduce discomfort and relieve stress during video calls?
RQ5. How could video calling tools be changed to better accommodate autistic users?

To answer these questions, we conducted 22 semi-structured interviews with autistic adults to learn about their perceptions of the benefits and drawbacks of VC. We found that they experienced difficulties with technology-mediated social norms at every stage of VC, from preparing for calls,

[1] While some in the autism community prefer people-first language, others have embraced the term "autistic" as their chosen identifying label, so we use that terminology as well [29].
[2] According to the neurodiversity framing of autism, there is a natural diversity of human neurological functioning, including autistic (neurodiverse) and non-autistic (neurotypical) cognitive styles [47].

VC can provide a rich experience because it offers a range of channels within one system, thus enabling users to leverage the affordances of multiple channels, e.g. video, audio, and/or text, among others. Media Richness theory helps explain how people choose CMC channels based on each channel's set of objective characteristics that determines its capacity to carry rich information [18]. In this context, information is rich if it helps the sender and receiver to communicate clearly and adjust their understanding as necessary. According to the theory, VC collaborators have a greater likelihood of conducting clear, unambiguous communications using audio and video than if they were communicating over a leaner channel (e.g. text) which would afford fewer sensory modalities. In addition to a channel's ability to carry rich information, users also chose channels based on how much control they have over them, their ability to remain anonymous if desired, and the ability to be co-present [17]. The notion of co-presence and distance is complicated by the collaborators' sense of distance; it is not as straightforward as whether or not the collaborators are literally in the same room. Research has looked at the effects of perception of distance and social context during VC. For example, Bradner and Mark found that VC collaborators exhibited more positive collaborative behaviors (e.g., cooperation) when they believed their collaborator was nearby, rather than far away [11].

Along with distance, other important attributes of the interaction partner that impact technology-mediated collaboration are social ties and the social factors relating to the context of the topic of conversation [17]. When text-based CMC became prominent, the social information processing (SIP) model [52] and hyper-personal model of CMC emphasized that users' experiences with CSCW were influenced by the social relationships among the participants. According to the hyper-personal model, users' social ties can become stronger through the coordination of four concurrent CMC *routines:* actions of the receivers, senders, channel, and continual feedback among those three components [53]. As these routines reinforce each other they can facilitate greater social desirability and intimacy that are of a different nature (and sometimes better) than developed in FtF interactions. This theory argues that attributes of CMC, such as temporal features, influence its capacity to build social relationships. For example, media preferences for populations that tend to face challenges of real-time FtF communication, such as people who are non-native speakers, often prefer text channels so that they can have more time to process messages and craft responses [46]. Although the SIP model and hyper-personal model focused on text channels, researchers have continued exploring social factors in CMC by extending the models to other modalities (e.g., online gaming [54]) and by conducting empirical research based on other CMC theories. The Embodied Social Proxy research explored a mobile VC terminal to enable interacting with a team in the context of social meeting spaces [51]. This approach evoked more social interactions that improved the remote collaborator's social integration with the team. How video technology is appropriated to reflect the context of social relationships can become more salient in a neurodiverse population that is sensitive to managing or controlling their sense of distance with their conversation partners. Our work is informed by this line of research, as we consider how autistic adults may experience strengthened social relationships using CMC channels that may be preferred over FtF interactions.

While VC naturally affords transmitting visual and nonverbal cues exchanged in an interaction, research has also begun to explore computationally detecting and representing those cues. Byun et al. [13] explored analyzing audio and video streams in a video call to detect visual and non-verbal cues and display them in real time to the interviewees to make their calls more successful. They found that users appreciated the feedback to help them manage their VC behavior. Grayson and Monk examined the establishment of mutual eye contact through VC, and found that users are able to learn to interpret eye gaze direction in VC, even without perfect mutual eye contact [25]. Beyond eye contact, facial expressions convey affective information and provide socio-emotional feedback. Facial expressions are particularly helpful in interpreting the affective state of users, illustrated

*even though it is awkward."* (P21). Focusing on their partner's face provided stimulus, making it *"a little bit easier to focus if I am having trouble that day."* (P17).

However, some interviewees found watching their partner's face to be quite overwhelming, similar to how they felt in FtF interactions. In fact, they revealed that they tried to *fake* eye contact by looking in between their partner's eyes, on their forehead, or in the general direction of their mouth and nose. P11 offered:

> *"I've had trouble with eye contact. What I used to do was focus on people's mouths because to me it was easier, and it made more sense to begin with because the mouth is the part that's moving. Over the years I've gotten better at it, and now I just focus on the center of their face." (P11)*

Interviewees commented on the fact that eye contact in a VC is never aligned, a situation that they actually appreciated:

> *"I probably make eye contact too much. But when you're video conferencing nobody knows where to put their eyes anyways. So everybody doesn't make eye contact. It's great. If I'm staring [or] needing to look off into space, it's not as apparent." (P06)*

Interviewees were often disturbed by other sensory signals transmitted by the video, such as bright lights. Fluorescent and undimmable lights in their physical environment, especially in spaces they could not fully control (e.g., their work office), were bothersome. Notably, a few interviewees said they experienced regular migraines due to bright lighting or the blue light emitted by computer monitors. They could also be bothered by the amount of light transmitted by their conversation partner's camera view especially in contrast to their own dimly lit environment.

Besides video, the interviewees also mentioned concerns about the audio in VC. Because they listened to their conversation partner's voice for changes in tone, volume, and intonation as cues to the meaning of what they said, they found background noises, such as cars, typing, or eating noises, to be distracting. These distractions often caused them to inadvertently switch their focus to the noise and lose track of the conversation.

In VCs with multiple people, interviewees noted that they became overwhelmed when people talked over each other or repeatedly interrupted each other. In some cases, interviewees responded by matching their conversation partners' styles just to get their points across. P08 described that he may not realize he was interrupting others, saying that *"it is entirely possible that I interrupt sometimes and don't notice, and it never comes to my attention that I just did that."*

*4.2.2 Cognitive Load.* Interviewees described many activities and aspects of interactions that required cognitive and emotional processing at a level that impacted their ability to maintain social interactions. These included topic familiarity, distractors, developing a mental model of their conversation partner, managing the conversation, emoting to others, reading other people's emotions, and reading other people's body language. Their description of managing the cognitive demands of VC illustrate the amount of effort involved.

Interviewees discussed how they cognitively processed input channels, often describing that they did best by focusing on one channel at a time. For example:

> *"Audio-wise, I can track what you're saying and going like that but as soon as I have to factor in body language, I can either pay attention to your body language or I can pay attention to what you're saying." (P03)*

Interviewees discussed feeling overwhelmed with the impromptu nature of VC, especially for work teams that keep their VC application running in the background. Their preference was to use email or talk FtF if the issue was urgent. P08 shared that he is *"notorious among my colleagues for never running [VC] in the background. You have to organize with me ahead of time for [VC]"* (P08). A

common reason among our interviewees for not keeping a VC application running is that they found it difficult to immediately switch to the topic of the impromptu VC or other CMC, such as texts and application notifications.

Cognitive processing appeared to remain an important part of their experiences, even after they felt they had acquired new socio-emotional skills. P14 described this learning experience as follows:

> *"Part of the autistic experience is learning ... basically most social interaction at work, there's a certain amount of manipulation to it, and pattern analysis. Like we actually have to consciously think about it a lot. But if you do that long enough, you start memorizing the patterns, and you find little shortcuts, so over time it gets a little bit easier for some people." (P14)*

In summary, a VC required meta-cognitive and low-level cognitive processing about the content of the meeting, technology-mediated interruptions from other people, and the surrounding social-emotional environment.

*4.2.3 Anxiety.* A consistent theme across all the interviewees' VC and FtF experiences was managing anxiety. They described many contributing factors to anxiety, including their role in the conversation, social familiarity, topic familiarity, conversational goals, adhering to social norms, and their current socio-emotional capacity. At a topic level, they expressed that they were more comfortable talking about concrete, familiar "things," such as games, technology, or work deliverables, more-so than when talking about "people." Situations that required negotiating, conveying nuance, or being unsure created stressful emotions. Interviewees discussed how their emotions, especially stress, in technology-mediated social interactions correlated with their relationships and the goal of the interaction. In a closer relationship, their stress was lowered. However, if the goal of the interaction was related to a conflict or ambiguous task, stress rose, even if they were socially close to the conversation partner. Our interviewees conveyed that their past emotionally-laden CMC conversations included: planning a vacation with others, collaborating and presenting in a design meeting, interviewing for a job, and talking to someone after they had a baby. Note that these are situations that can heighten emotions for even neurotypical people.

Anxiety drove some interviewees to staunchly *avoid* using video web cameras, two said they did not even own a camera. These interviewees described feeling self-conscious about being on camera and felt judged by others. They felt less freedom to move around, multi-task, stim, and fidget. In addition, the experience of seeing themselves in the video camera was uncomfortable. P15 described it *"like staring at a giant mirror for an hour."*

During VCs, the interviewees strove to adhere to neurotypical social norms, such as eye contact, as described in the section above about Sensory Sensitivities. Since the way they expressed themselves verbally and through body language sometimes differed from those social norms, they had anxieties about being misunderstood, as P19 described:

> *"Someone may think that I'm either seeming upset if I'm thinking about something, an autistic person may feel like that. Or if I'm trying to say something in a different way, sometimes my voice tone may sound disrespectful when it meant to be respectful." (P19)*

They also worried about misreading their conversation partner because they would miss nonverbal cues such as *"the tightening of the face, a straining of the smile. Things like those I may not see unless I am actively looking for them."* (P19)

Interviewees described feeling significantly less anxious when talking to other autistic people. They felt more comfortable to be themselves because the social norms were more intuitive and they did not have to mask their autism. They could bond over common expectations about their

dynamics by constructing a set of social meeting rules to be agreed upon by participants in advance. One example of such a rule was strict time-boxing, where participants who ran over an appropriate contribution window were muted by the autistic organizer:

> *"During the meetings, we allot each person a certain amount of time to speak. We let them know, and if we try to tell the person to stop and they won't, we mute them. It's a bit over-controlling but the members understand in the end that we have a limited amount of time for chapter meetings. Other people need to be heard, and it is not all about them, and especially if they may go off on tangents maybe they want to talk about school or something but it's not relevant. You can mute them individually." (P19)*

Some autistic people admitted to extensively researching future VC group participants online to help them formulate a mental model of other VC participants, leading to greater prediction of actions and confidence approaching those members during the call. Interviewees liked to turn their camera and microphone off before the start of the meeting, however acknowledging that these adjustments can cause increased anxiety when it comes time to contribute: *"I scramble to turn the microphone back on."* (P13) Interviewees were careful to hide personal belongings from view, and an emphasis on personal appearance and hygiene increased with formality of the call. Some interviewees who were more familiar with VC experienced less anxiety, and therefore made fewer preparations, especially with personal calls.

*4.3.2 Coping Strategies during Initiation Phase.* As the VC begins, there are continued sources of sensory, cognitive, and social concerns for our interviewees, all of which need to be managed. In general, they feel anxious about the initial *"two–three minutes of ... chaotic mess"* (P09), which includes disorganization as other VC participants join, and the pressure to adhere to awkward social pleasantries, namely, small talk. Interviewees expressed feeling nervous anticipation, and sometimes irritation, about audio, video, and connectivity issues with VC. There seems to be an exception for VC with close friends, in which technical issues are less of a concern. When entering a VC, many interviewees felt the immediate spotlight was uncomfortable. Here, P15 explains the contrast between in-person conversation and VC in regard to immediacy:

> *"[In person,] I won't know who's in a room [before I enter] ... but once I walk into the room, then I see people before I actually have to walk up to them and talk to them, you know? So at least I have a couple seconds warning. Whereas with video chat, boom I'm there." (P15)*

The majority of sensory concerns during the initiation phase are visual distractions and unhelpful pixelated visuals. P03 experiences the following camera anxiety:

> *"The new [VC] thing that I don't like is when you launch it, I always have ... anxiety on whether it's going to turn on my camera or not. Sometimes it does turn on the camera/turn off the camera. The first couple times it launched and turned on a camera, I was like, oh, this is no good." (P03)*

Some of this was controlled for by disabling the video prior to connecting. However, interviewees noted this can later cause confusion when it comes time to present their desktop. We saw similar feelings around microphone disabling and enabling during initiation.

Presenting is especially burdensome due to potential technical hiccups, such as poor bandwidth or a weak or lagging wireless connection. Such technical concerns consume cognitive resources at the time of VC initiation. General questions at the outset of a call include *"Are we on," "Can you hear me,"* and *"Can you see me"*. (P03) For this reason, the *"beginning is especially awkward."* (P03) These are sentiments that can likely be shared by all VC users; however, the effects appear to be amplified for the autistic segment. P07 reported stress for hours prior to a call, and fusses with her camera

for 30 minutes before initiation, to reduce awkwardness and mitigate technical complication. P15 also explains the discomfort of initiating VC:

> *" ... it's really awkward for the first five or ten minutes or so. And then, usually it becomes better as I sort of get used to the- I don't know how to say it. Just, I guess get acclimated maybe to the video call and all of that." (P15)*

Social coping during VC initiation is a natural continuation of the coping strategies employed in VC preparation: autistic adults often strive toward mental model completeness of their VC participant(s). In the initiation phase of VC, they do this via the avenues previously mentioned: small-talk, introductions, and role-establishment, much of which is an added source of pressure and anxiety. An interrupted VC initiation causes anxiety by reducing the autistic adult's sense of control and certainty, a natural side-effect of the then hampered predictability of their interaction: *"[missing introductions] puts me a disadvantage"* (P13). By creating a mental rendition of their conversation partner(s), autistic adults in VC can better calculate the optimal interaction style for each participant, which is consistent with the hyper-systematizing previously mentioned.

Autistic users are highly considerate and worried about the message being portrayed by their facial expression. Many are careful to plaster a smile on their face out of concern that their natural facial expressions may mislead the receiver into thinking that they are disliked. Some interviewees found the video preview very helpful because it showed them exactly what their conversation partner was seeing. P17 used her video preview to monitor her facial expressions. Many interviewees discussed that the misalignment of the camera views between them and their partner allowed them to *"fake eye contact by looking at my web camera, so the person thinks I'm looking at them, but I don't have to feel them looking at me"*. (P14) Note that these attempts to avoid discomfort, and the sending of incorrect conversational signals, is a highly conscious effort.

*4.3.3 Coping Strategies during Participation Phase.* After the initiation phase, the VC discussion shifts into the body of the meeting or the main topic or activity of a social VC. During the participation phase, the interviewees needed to actively moderate sensory inputs. Interviewees who perceive a lot of information from the tone of someone's voice desire a clear and consistent voice channel. They actively manage audio in the VC system to adjust speaker volumes—individually for each speaker if that was supported in their VC system. They adjusted the light coming from their partner's web cam to minimize painful bright lights. They engaged in repetitive behaviors like fidgeting, stimming, and walking around, which helped them release energy and focus on the content of the meeting. If they were running a meeting with people they knew well, they advised them about how to minimize sensory distractions. For example, for a monthly meeting, P19 told remote VC interviewees to stay in a quiet place, since otherwise background noise *"disrupts the entire meeting, and so we'd have to mute them"*.

Interviewees had a specific set of coping strategies to manage their attention during a VC. To focus on the content of the conversation, many interviewees turned off their own camera and the streaming video of their partner. By reducing their visual inputs, especially that of another person, they could better concentrate, as described by P09, *"if I'm trying to concentrate, then, I stare at my phone, and I can really absorb what people are saying, or I take notes at the same time."* A benefit of turning off their own camera was that they could multi-task and felt more comfortable engaging in repetitive behaviors. Interviewees could become distracted by the video stream of their partner when there were background movements or actions of the meeting participants, especially if they were doing something repetitively like typing. If a conversation partner was sharing their desktop, their mouse cursor movement could be distracting. Some interviewees stressed that they were audio learners and relied on their auditory skills to read the tone of the speaker's voice, and therefore, felt they were more effective communication partners with the video stream off.

Interviewees were actively engaged in meta-cognition tasks during the participation phase. Several interviewees described ways that they externalized knowledge to help them process the content of the meeting. Their strategies included taking notes, looking at off-line copies of presentation slides, or reading the VC chat window—rather than focusing on the verbal discussion. Some interviewees had difficulty tracking topics, sometimes because they were still thinking through a point of a previous topic or formulating their thoughts because they wanted to express them thoroughly. Tracking topics was also difficult due to notifications from other applications or the VC system, which caused them to context switch. In those cases where the VC connected with a conference room with multiple participants, our interviewees experienced stress and cognitive load trying to identify who was talking.

There were some scenarios in which interviewees were concentrating on one topic and did not seamlessly progress to the next topic of discussion. For example, if they were discussing a topic that they were very interested in, they could become so engrossed that they could speak at length. Due to their excitement, they sometimes did not pick up on their partner's nonverbal cues about wanting to chime in or change the subject. In the scenario of listening to a VC, interviewees may be contemplating a point from one slide, not noticing that the presenter has moved onto the next slide.

In terms of coping strategies to support social relationships, interviewees were conscious of how they were being perceived by others. On one hand, a few interviewees expressed that they like how they come across in VC, such as P03 who said, *"virtual me is better and more dynamic than in-person me."* However, the majority of interviewees worried that they would be misunderstood or harshly judged over VC. The VC video preview was a useful tool to check in on their facial expressions and their body position.

> *"You can kind of know exactly what the person is seeing as well which is sort of unique because I can see, there's a preview at the bottom where I know exactly what's going across to the other person and that's very helpful for me." (P17)*

A surprising source of social anxiety was the usability difficulties of the VC application beyond technical connectivity issues. One interviewee described at length how anxious she got sharing her desktop, knowing her computer actions were being observed. She worried that her teammates would judge her for experiencing usability issues such as accidentally clicking on the wrong button or being unable to locate the "stop sharing" button:

> *"It is uncomfortable using my desktop incorrectly while I'm sharing desktop. I always forget where the share button is... Watching someone else do it, it seems so easy." (P13)*

Interestingly, interviewees had different perspectives on what constituted the efficient use of time during VCs, mostly depending on the purpose of the VC. For work related VC, interviewees generally perceived that VC conversations were more concise than they are in FtF situations in which people can talk for unpredictable amounts of time and meander into unanticipated topics. On the other hand, interviewees used VCs to hang out with people they are close with and with fellow students for online study dates. In these situations, they appreciated being able to take pauses and take advantage of being in their own homes to take care of other tasks or lie down on the bed to relax before resuming the VC.

*4.3.4 Coping Strategies during Termination Phase.* Closing out a VC presents yet further challenges for autistic interviewees. In terms of sensory sensitivities, the main issue reported in this phase was that interviewees would "run out of their spoons" before the end of the actual conversation. This can lead to their abruptly leaving the call. This cognitive load exerted in this phase also extends past the end of the actual VC. Interviewees reported being anxious about ambiguous action items

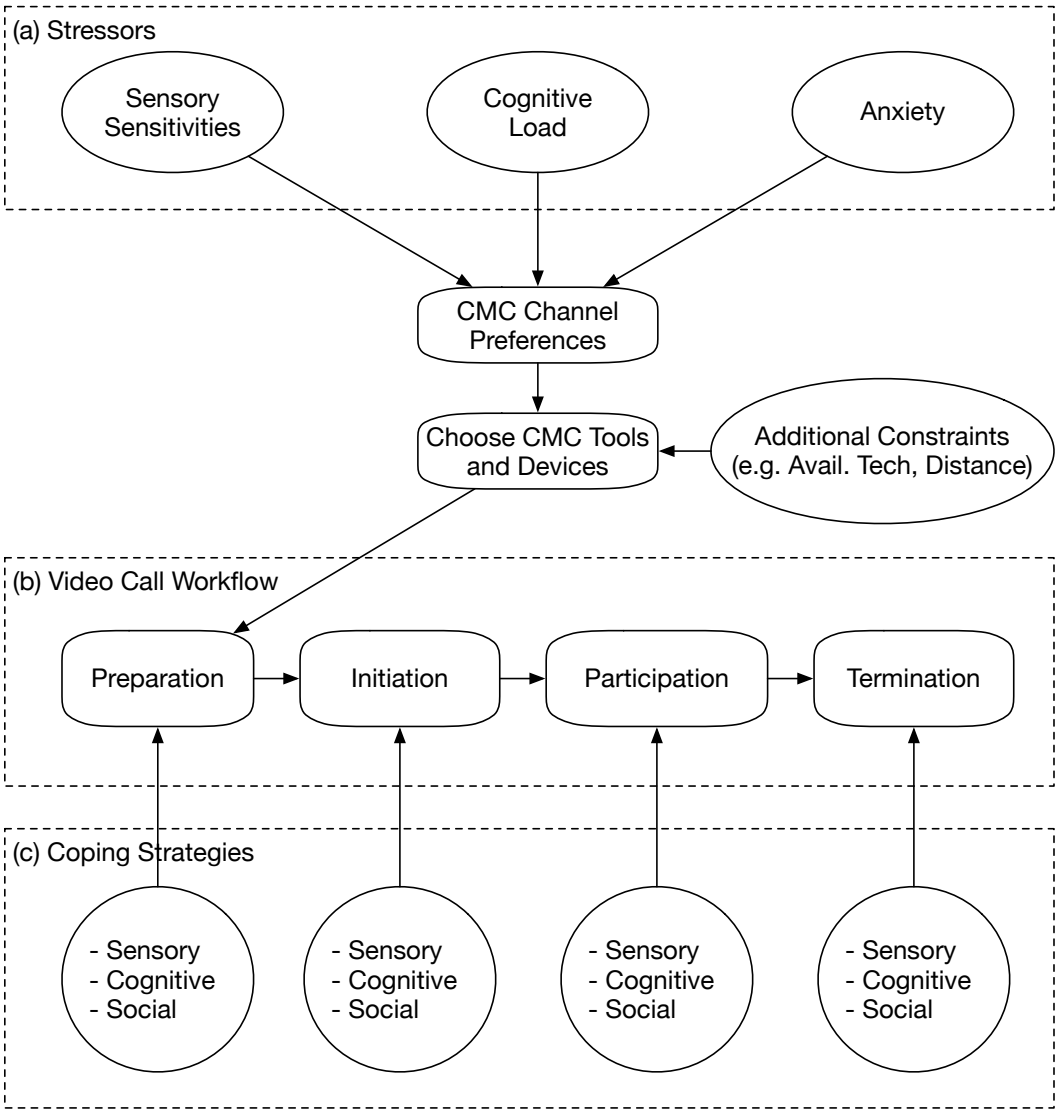


Fig. 1. A neurodiversity-sensitive computer-mediated communication model (NDS-CMC): (a) Anxiety, cognitive load, and sensory sensitivity are stressors that affect autistic users. These factors influence which CMC channels are used in a video call, in additional to the usual constraints. Once decided, autistic users invest in a variety of (c) sensory, cognitive, and social coping strategies to handle each of the (b) 4 stages of the video call workflow.

can help us understand their effects. We then expand on several of the more interesting findings and relate them to theory and prior literature to help explain their presence and their impact on CMC. Finally, we discuss how conversations among people on the autism spectrum mediate factors that affect the model.

asynchronous CMC channels (e.g., text or email) to minimize the need to perform tasks that were cognitively-demanding for them. Scholl et al. found that people preferred text-based chat over audio because it was less intrusive, and was easier to use, especially for communicating in a second language, something that incurs more cognitive load than using a native language [45]. As autistic people "use up spoons," their socio-emotional cognitive resources decline, making tasks even more difficult and sometimes leading them to abandon the conversation. Incidentally, VC's inability to support direct eye contact is actually an *advantage* for most autistic people, who find it cognitively draining to do this when speaking to someone FtF.

Interviewees described a complex dynamic between their social-emotional skills and their agency in the conversation. In terms of expressing their own emotions, their emotional affect did not always come naturally; rather, they had to make a conscious effort to emote both positive and negative emotions. Their lack of facial expression was sometimes interpreted by their neurotypical conversation partners as anger, even though they were feeling happy. Since this negatively affected the conversation, autistic adults would spend considerable cognitive effort to deliberately make facial expressions just for the benefit of their neurotypical partner. Their capacity to emote was impacted by their sensory comfort levels, their energy levels, and their relationship with their conversation partners. One way to conserve energy was to rapidly move their eyes between several people in a conversation, minimizing the time spent looking at any one person's eyes but still believing that the others felt they were looking at them.

Hyper-systemizing theory helps explain the cognitive difficulties experienced by autistic adults when trying to keep up with the flow of a conversation, especially an inefficient one. The flood of (potentially irrelevant) details in a conversation requires intense cognitive effort to manage; their ability to process input in larger, more abstract chunks is limited. When trying to read others' social-emotional cues, our interviewees intentionally scanned for nonverbal cues—something that was difficult for them and which lowered their focus on the verbal conversation. Their lack of a theory of mind explains their difficulties in reading the emotions and body language of other people. However, we observed something more. Interviewees reported having an easier time reading positive emotions and body language (such as happiness and laughter) than recognizing anger or sadness. This may be because people tend to suppress negative emotion expression in social interaction [26, 50]. They said it was especially difficult to read emotional cues when the sender's words where incongruent with their tone of voice and body language. This occurs when someone is being sarcastic or hiding their true feelings.

Interviewees also told us of the need to mask their autistic behaviors to conform to neurotypical politeness norms, even though it required intense cognitive effort. Tannen showed that when conversing, neurotypical people try to match their *conversational style*, a concept that includes vocal pace, prosody, and relative volume [49]. For the autistic partner who has trouble perceiving these vocal characteristics, it would be difficult to please their partner by matching them. A similar challenge occurs with misinterpreted and mismatched facial expressions and body language. Several interviewees reported that their neurotypical conversation partners would often misinterpret their lack of facial expression as anger, even though they were feeling happy.

Interviewees told us that the stressors that induce anxiety were a constant worry for them, and influenced their CMC preferences. Only when interviewees were familiar with their conversation partners and the topics of conversation did they report feeling most comfortable in an in-person or high-bandwidth video call. Otherwise, they made use of low-bandwidth CMC channels like texting or email to keep conversation partners at a distance, relieving their anxiety over the impending situation. This fits with a theory of mind and the hyper-systemizing theory, which explain that when having difficulty anticipating the responses of a conversation partner, autistic adults will fear that saying the wrong thing will cause the conversation to break down.

Banff, Alberta, Canada, 323. https://doi.org/10.1145/1180875.1180925

[46] Leslie D. Setlock and Susan R. Fussell. 2010. What's it worth to you?: the costs and affordances of CMC tools to asian and american users. In *Proceedings of the 2010 ACM conference on Computer supported cooperative work - CSCW '10*. ACM Press, Savannah, Georgia, USA, 341. https://doi.org/10.1145/1718918.1718979

[47] Judy Singer. 1999. 'Why can't you be normal for once in your life?' From a 'problem with no name' to the emergence of a new category of difference. In *Disability Discourse*, Judy Singer and Sally French (Eds.). Open University Press, Philadelphia, PA, 59–67.

[48] Daniel P. Skorich, Tahlia B. Gash, Katie L. Stalker, Lidan Zheng, and S. Alexander Haslam. 2017. Exploring the Cognitive Foundations of the Shared Attention Mechanism: Evidence for a Relationship Between Self-Categorization and Shared Attention Across the Autism Spectrum. *Journal of Autism and Developmental Disorders* 47, 5 (01 May 2017), 1341–1353. https://doi.org/10.1007/s10803-017-3049-9

[49] D. Tannen. 2005. *Conversational Style: Analyzing Talk Among Friends*. Oxford University Press, New York.

[50] Silvan S Tomkins. 1984. Affect theory. In *Approaches to emotion*, Klaus Sherer and Paul Ekman (Eds.). L. Erlbaum Associates, Hillsdale, NJ, 163–195.

[51] Gina Venolia, John Tang, Ruy Cervantes, Sara Bly, George Robertson, Bongshin Lee, and Kori Inkpen. 2010. Embodied social proxy: mediating interpersonal connection in hub-and-satellite teams. In *Proceedings of the 28th international conference on Human factors in computing systems - CHI '10*. ACM Press, Atlanta, Georgia, USA, 1049. https://doi.org/10.1145/1753326.1753482

[52] Joseph B. Walther. 1992. Interpersonal Effects in Computer-Mediated Interaction: A Relational Perspective. *Communication Research* 19, 1 (1992), 52–90. https://doi.org/10.1177/009365092019001003 arXiv:https://doi.org/10.1177/009365092019001003

[53] Joseph B. Walther. 1996. Computer-mediated communication: Impersonal, interpersonal, and hyperpersonal interaction. *Communication Research* 23, 1 (1996), 3–43.

[54] Joseph B. Walther, Brandon Van Der Heide, Artemio Ramirez, Judee K. Burgoon, and Jorge Peña. 2015. Interpersonal and Hyperpersonal Dimensions of Computer-Mediated Communication. In *The Handbook of the Psychology of Communication Technology*, S. Shyam Sundar (Ed.). John Wiley & Sons, Ltd, Chichester, UK, 1–22. https://doi.org/10.1002/9781118426456.ch1

[55] Peter Washington, Dennis Wall, Catalin Voss, Aaron Kline, Nick Haber, Jena Daniels, Azar Fazel, Titas De, Carl Feinstein, and Terry Winograd. 2017. SuperpowerGlass: A Wearable Aid for the At-Home Therapy of Children with Autism. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 1, 3 (Sept. 2017), 1–22. https://doi.org/10.1145/3130977

[56] Jurgen Wegge. 2006. Communication via Videoconference: Emotional and Cognitive Consequences of Affective Personality Dispositions, Seeing One's Own Picture, and Disturbing Events. *Human-Computer Interaction* 21, 3 (Sept. 2006), 273–318. https://doi.org/10.1207/s15327051hci2103_1

[57] Heinz Wimmer and Josef Perner. 1983. Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition* 13, 1 (1983), 103–128.

[58] Annuska Zolyomi, Anne Spencer Ross, Arpita Bhattacharya, Lauren Milne, and Sean A. Munson. 2018. Values, Identity, and Social Translucence: Neurodiverse Student Teams in Higher Education. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems - CHI '18*. ACM Press, Montreal QC, Canada, 1–13. https://doi.org/10.1145/3173574.3174073